



TITLE:

Development of computational analysis tools for natural products research and metabolomics( Abstract\_要旨 )

AUTHOR(S):

Ahmed, Mohamed Fathi Youssef Mohamed

---

CITATION:

Ahmed, Mohamed Fathi Youssef Mohamed. Development of computational analysis tools for natural products research and metabolomics. 京都大学, 2016, 博士(薬科学)

ISSUE DATE:

2016-03-23

URL:

<https://doi.org/10.14989/doctor.k19673>

RIGHT:

許諾条件により本文は2017-03-22に公開

( 続紙 1 )

|  |   |    |                       |
|--|---|----|-----------------------|
| 京都大学   | 博士（薬科学）   | 氏名 | Ahmed Mohamed Mohamed |
| 論文題目   | Development of computational analysis tools for natural products research and metabolomics<br>(天然物科学およびメタボロミクスのための計算解析ツールの開発) |    |                       |
| <p>Metabolic analysis in living organisms is important for understanding biological systems, having wide applications ranging from therapeutics, drug discovery and biotechnology. For example, metabolic profiles can be used to identify biomarkers for early disease prognosis. In natural products research, bioactive secondary metabolites are considered as new drug leads. In biotechnology, metabolic engineering is routinely used for optimization of metabolite production.</p> <p>Depending on applications, metabolic analysis can be carried out by one of two paradigms: network analysis and metabolite identification. Firstly, network analysis investigates metabolic networks to systematically identify active metabolic pathways and metabolite production patterns. So, metabolic network analysis is used for biomarker discovery and metabolite production optimization. Secondly, in metabolite identification paradigm, the presence and concentration of individual metabolites are investigated. For example, discovery of new drug leads from natural products involves structure determination of individual metabolites with promising bioactivities or novel chemical scaffolds.</p> <p>Despite the importance of metabolic analysis, necessary computational tools are still lacking. The technological advances increased the amount of experimental data that can be collected, making manual analysis challenging. For example, analysis of genome-scale metabolic networks with thousands of metabolites in manually infeasible, requiring computational tools. In natural products research, integration of computational tools with spectral databases are needed for rapid identification of known compounds. Also, software tools for online processing of NMR measurements, a central technique for metabolite identification, are still lacking. Easy-to-use computational tools enable researches to quickly analyze and interpret experimental data, reducing cost and effort.</p> <p>In this thesis, we explore computational methods and tools needed for different paradigms for metabolic analysis, presenting two novel tools, NetPathMiner and NMRPro. First, we present NetPathMiner, a software in R framework, for identification of active metabolic pathways based on gene expression. Second, we review computational resources for rapid</p> |   |    |                       |

identification of natural products identifying the need for software tools for processing nuclear magnetic resonance (NMR) spectra. Finally, we present NMRPro, a web component for online interactive processing of NMR spectra. We discuss each topic briefly below.

NetPathMiner is a general framework for mining, from genome-scale networks, paths that are related to specific experimental conditions. NetPathMiner interfaces with various input formats including KGML, SBML and BioPAX files and allows manipulation of networks in three different forms: metabolic, reaction and gene representations. NetPathMiner ranks active paths and applies clustering and classification to the ranked paths for easy interpretation, providing static and interactive visualizations of networks and paths.

Rapid identification of previously isolated compounds in an automated manner, called dereplication, steers researchers toward novel findings, thereby reducing the time and effort for identifying new drug leads. Dereplication identifies compounds by comparing processed experimental data with those of known compounds, and so, diverse computational resources such as databases and tools to process and compare compound data are necessary. Automating the dereplication process through the integration of computational resources has always been an aspired goal of natural product research. To increase the utilization of current computational resources for natural products, we provided an overview of the dereplication process, and then listed useful resources, categorizing into databases, methods and software tools and further explained them from a dereplication perspective. Finally, we discussed the current challenges to automating dereplication and proposed solutions.

Finally, we present NMRPro, an integrated web component for interactive processing and visualization of NMR spectra. Web applications are well used recently because they are platform-independent and easy to extend through reusable web components. Although available web applications can analyze NMR spectra, they still lack the necessary processing and interactive visualization functionalities. Incorporating NMRPro into current web applications enables easy-to-use online interactive processing and visualization.

In conclusion, we surveyed the current status of computational tools for metabolic analysis and presented two novel tools, which can be building blocks for automating research in natural products and metabolomics.

(論文審査の結果の要旨)

メタボロミクス等、化合物のハイスループットデータが大量に得られる時代を迎えている。しかし、このような大規模実験データを効率的に処理できるソフトウェアが十分に整備されているとは言い難い。特に、短時間で計算できる解析を装備し、解析結果を視覚的に把握でき、また、Rパッケージ等ウェブ上で簡単に取得できる簡便なツールはこれまで無かった。本研究は、それらを実装しようとする試みである。

具体的には以下の2つの内容からなっている：

1、様々なファイルフォーマットの代謝パスウェイデータベースを入力とし、ネットワーク解析結果を視覚的に表示するRパッケージの開発。

2、天然物化合物のNMRデータの解析のためのウェブツールの開発。この研究開発に対しては、同様のソフトウェアの現状に対する幅広いサーベイが付け加えられている。

このようなソフトウェアのアカデミックの研究開発において、現在の情報科学技術の進展からすれば、機械学習等知的な情報処理技術やアルゴリズムの計算効率性等も可能なあるいは主要な研究対象となるべきであろう。本研究はそのような情報科学における先端的な研究を含んではいない。また、ツールを使うことによって得られるであろう、生命科学の現象への深い洞察を含んでいるわけではない。さらに、将来に渡って、本来、完成されるべきであろう、そのようなツールの全体像に較べれば、最初のほんのわずかな一步を記したにしか過ぎないソフトウェアである。しかし、現状では得難いにも関わらず必要とされている、生命科学の研究を推進するための補助的かつ不可欠なツールの開発であり、これらのツールの誕生により二次的に生命科学の研究が大きく推進されることが期待される。また、これらのツールは化合物、特に天然物化学からの独自の着眼点を生かしたものであり、生命情報科学や関連領域においても類を見ない、独創性がある研究開発と認められる。

さらに、開発されたツールとサーベイは、生命情報科学の代表的な雑誌であるBioinformatics誌に2篇、Briefings in Bioinformatics誌に1篇、計3報として掲載されており、生命情報科学の研究成果としての一定の基準を満たしているとみなすことができる。

よって、本論文は博士（薬科学）の学位論文として価値あるものと認める。また、平成28年2月26日、論文内容とそれに関連した事項について試問を行った結果、合格と認めた。